DOCUMENT RESUME

ED 338 242 IR 053 752

AUTHOR Culmer, Carita M.
TITLE Converting Inhouse Subject Card Files to Electronic Keyword Files.
PUB DATE 91
NOTE 54p.; Practicum Report, Nova University.
PUB TYPE Dissertations/Theses - Practicum Papers (043)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Libraries; *Access to Information; *Bibliographic Databases; Card Catalogs; Community Colleges; Controversial Issues (Course Content); Hypermedia; Indexes; *Information Retrieval; Library Automation; Library Materials; Library Technical Processes; Microcomputers; Online Catalogs; *Online Searching; Search Strategies; Subject Index Terms; Two Year Colleges; User Needs (Information)
IDENTIFIERS HyperCard; Keywords

ABSTRACT

The library at Phoenix College developed the Controversial Issues Files (CIF), a "home made" card file containing references pertinent to specific ongoing assignments. Although the CIF had proven itself to be an excellent resource tool for beginning researchers, it was cumbersome to maintain in the card format, and was limited to very general subject headings. The goal of this project was to make specific information in the CIF more readily accessible to students using a method of file maintenance that would require less library staff time to input new records or alter old ones. To this end, an electronic format was developed for CIF using HyperCard 2.0 on a Macintosh SE. The conversion procedure included experimentation with various formats, the creation of electronic templates, the development of input procedures and creation of a procedures manual, training librarians in search techniques, training and supervision of support staff who would do the actual inputting of data, and proofreading. A core sample demonstrated that bibliographic records of in-house files could be enhanced, and retrievability improved, by including full tables of content rather than a keyword index or Library of Congress subject headings. Real time experiments indicated that data input could be accomplished in less time than hand production and filing of cards. (Reports on data input experiments are appended to this paper.) Instructions for students to use the file are simple and basically the same as for searching keywords on the online catalog (OPAC). Once conversion has been completed, copies of CIF on floppy diskettes will be made available to other libraries in the community college district. It is noted that, although problems had been encountered in using OCR scanning for input, this process was still under consideration, and that searching capabilities would be improved if HyperCard could be programmed to handle Boolean logic. (MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED338242

Converting Inhouse Subject Card Files to Electronic Keyword Files

A Practicum Report by

Carita M. Culmer
Nova University
Fort Lauderdale, FL

IR053752

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carita Culmer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Abstract

Converting Inhouse Subject Card Files to Electronic Keyword Files.
Culmer, Carita M., 1991: Practicum Report, Nova University,
The Center for Advancement of Education.
Descriptors: Academic Libraries/ Access to Information/ Databases/ Indexes/ Information Retrieval/ Information Systems/ Libraries/ Research Libraries/ Library Automation/ Library Catalogs/ Library Materials/ Library Technical Processes/ Selective Dissemination of Information/ User Needs (Information)/

Academic libraries develop "home made" card files containing references pertinent to specific ongoing assignments. With the advent of electronic databases, information retrieval has been expanded on a massive scale. Similar research databases can be miniaturized for inhouse files. A program was develped, using HyperCard 2.0 on a Macintosh SE. A core sample demonstrated that bibliographic records of inhouse files could be enhanced, and retrievability improved by including full tables of content. Real time experiments indicated that data input could be accomplished in less time than hand production and filing of cards.

Appendix S

Authorship Statement

I hereby testify that this paper and the work it reports are entirely my own. When it has been necessary to draw from the work of others, published or unpublished, I have acknowledged such work in accordance with accepted scholarly and editorial practice. I give this testimony freely, out of respect for the scholarship of other professionals in the field and in the hope that my own work, presented here, will earn similar respect.

Signed: Carita M. Culmer
Student's name

Minor contributions of a professional character and comparable nonprofessional assistance may also be acknowledged.

Table of Contents

Page

Title page i

Abstract ii

Authorship Statement iii

Observer's Verification iv

Table of Contents v

Chapters

I. Purpose 1

II. Research and Solution Strategy 11

III. Method 19

IV. Results 35

V. Recommendations 40

References Cited 42

Appendices

Appendix A: Time required to establish keywords 44

Appendix B: Controversial Issues File Survey .. 46

Appendix C: Time Required to Assign Subject Headings, Produce & File Cards After Subject Headings Have Been Assigned..... 47

Appendix D: Time Required to Enter Data, Inclucing Full Table of Contents 48

Appendix E: Time Required to Underline Keywords and Proofread Electronic Records 49

CHAPTER I

Purpose

Background.

The college, established in 1920, is the parent institution of a ten-campus community college district. It is a non-residential campus, the majority of the students being residents of nearby communities. The college offers both academic transfer and vocational programs. Fall 1990 enrollment was 12,930 students. Total district enrollment was 91,202 students. Enrollment figures are 45th day certified. The student ethnic population for Spring 1990 included 2.7% American Indian, 2.5% Asian, 4.8% Black, 13.3% Hispanic, and 76.7% White or other designation (Day, 1990). The enrollment reflects the multi-ethnic, multi-cultural population of the metropolitan community.

The college library houses 83,360 books, 11,212 bound periodicals, 205 periodical titles in microform, and subscribes to 527 current periodicals. For the last fiscal year, 209,000 people made use of the library, most of whom were students. During this time, 62,000 books and 14,113 periodicals were checked out. (These figures do not reflect books and periodicals

used in the library but not checked out.) Five full-time and six part-time librarians conducted 154 classroom bibliographic instruction sessions, and assisted 19,000 library users needing help with assignments or other academic matters (Annual Report, 1989/90). The author has been a full-time librarian in the college library for the past twenty years.

Academic libraries exist primarily to cooperate with teaching faculty and help students with their assignments. It was for this reason that the Controversial Issues File (CIF) was created.

Community college students are equivalent to freshmen and sophomores in a four-year college or university. They are at the very beginning of their academic pursuits, and need help in developing research skills. Research assignments might include term papers, essays, panel discussions, debates or pesuasive speeches for English, Speech, or other classes. Such assignments often require that the student take a stand on one side or the other of a current controversial issue.

Periodical articles usually focus on one side of an issue, as do many books. However, assignments often require that students explore literature which analyzes

both sides of an issue, rather than the more common polemic position of most publications. Some periodical publications, and some book series specialize in presenting arguments on both sides of hotly debated topics. The beginning researcher can make a very good start by using one or more of these publications. Each follows a somewhat different format, but all are solidly structured, clearly presented, well-balanced and reliable.

Until December 1987 the college library had a card catalog and numerous printed periodical indexes. In January 1988, the library went online with an automated library system which includes a District-wide online public access catalog (OPAC). While the old card catalog listed only the holdings of the campus library, the OPAC contains the holdings of all campus libraries throughout the District.

At nearly the same time, the library subscribed to an online periodical index. Unlike printed indexes, which are issued periodically throughout the year and cumulated annually, this index is completely updated every month, includes three plus years of article citations, and covers approximately twice as many magazines as an equivalent index printed monthly.

While the OPAC greatly increased the total number of resources which a student could locate at any one time, it created a corresponding burden of that many more records which the researcher had to examine to determine applicability to the subject at hand.

Similarly, the online index, being very broad-based and cumulated over an extended period of time, confronted the student with an almost overwhelming list of articles on the chosen topic. Each bibliographic entry had to be examined to determine its usefulness for the assigned area of research. To add value to this index, a combined list of periodical holdings on all campuses encouraged the student to consider even articles in publications not owned by the home campus.

Since the introduction of OPAC and the online index, other expanded indexes have been added, both online and on compact disc - read only memory (CD-ROM). Keyword searching is now a part of OPAC, as it is for the CD-ROM indexes. Every new technological advance brings with it greatly increased numbers of bibliographic records on every conceivable subject. Experienced researchers can locate pertinent records fairly efficiently, but the beginning researcher is sometimes overwhelmed by the sheer volume of records.

Thus, early in 1988, the Controversial Issues File (CIF) was created. The process began with the librarians discussing known periodicals or series which frequently or regularly presented scholarly treatments of both sides of debatable social issues. All librarians spend a portion of each working day helping students locate appropriate library materials. Selections were based on experience and a working knowledge of the chosen publications.

The purpose of creating a separate file of limited proportions on a relatively small number of topics was twofold. First, it would give the beginner a good start in the research process. Instead of being faced with a large number of possibilities, and having to work through many which were not applicable because they did not present balanced, analytical coverage, a smaller list assurred some measure of success and encouragement very quickly. Second, the chosen publications are well structured in ways which help the beginner to formulate progression and development of the topic in whichever direction was assigned. It should be noted here that students who are introduced to the CIF are also advised strongly that this is only a starting point; that, once sufficient material is

gleaned to form a workable focus and outline, other indexes and the OPAC must be used to locate additional material which will contribute to a college-level presentation.

Initially, the CIF consisted of printed lists, and photocopies of tables of content and individual indexes, all of which were soon replaced by a card file, in alphabetical order by subject. The file is limited to five subscription titles -- two periodicals and three monograph series. The five publications are Congressional Digest and Editorial Research Reports, both of which are periodicals, and Opposing Viewpoints Series, Reference Shelf Series, and Taking Sides Series, all of which are books.

Congressional Digest is a monthly publication, and Editorial Research Reports is published weekly. Both of these periodicals devote each entire issue to a single topic. Each Opposing Viewpoints book is a collection of articles presenting opposing points of view on the many facets of complex social concerns. Reference Shelf books contain background information, articles, excerpts from books, and speeches on debatable social issues. Taking Sides books present one "pro" and one "contra" article on each of as many as

two dozen points of a single issue, plus a summary and a postscript to each. Writings in each of these five sources are by recognized authorities and well-known authors.

Cards for the weekly periodical Editorial Research Reports are hand-written; entries for the other four publications are computer-generated using a Quick Card program. Each entry is very basic, containing only a standard Library of Congress (LC) subject heading, article or book title, periodical or series title, date, and page numbers. In addition, the cards for the three series contain content notes and library call numbers.

Problem Statement.

The CIF had proven itself an excellent resource tool for beginning researchers. However, it was cumbersome to maintain in its present format, and was limited to the confines of subject headings which conveyed only the broadest general idea of any one item's contents. This was not the fault of LC subject headings which, individually, were very precise. Rather, it was a matter of economy. Records which were hand produced, and which required a separate card for each subject heading, took a considerable amount of time to produce and file individually. What was needed

was a method of producing records with greater detail, but without increasing the workload of an already overburdened staff.

Outcome Objectives.

Ultimately, the goal is to make the completed electronic file available to other campuses. The five resources which make up the CIF are standard academic library purchases, available in nearly every community college library. Students on campuses both in and out of the District could benefit equally from using this new technology to access information.

The goal of this project was to make a good service, the Controversial Issues File (CIF), much better. A definition of "better" in this case was a method of file maintenance which a.) made specific information readily accessible to students, and b.) required less, or at least no more, library staff time to input new records or alter old records, thus assurring that the most current information became available to students in the shortest possible time frame. In its card file format, the CIF made information available only through broad subject headings. By introducing the mechanism to search keywords, students gained the ability to determine

whether specific aspects of a topic were contained in CIF sources.

Keyword searching does require brainstorming on the part of the student. The student must give the matter some thought and think of synonyms or alternative approaches. Another way is to begin with a broad subject, such as GUN CONTROL, enter that as an electronic search, then read the contents information for clues to narrowing the topic. Either way, the student gets a good start on learning research techniques.

The process quickly eases the novice into a critical thinking mode. Helping students to develop their own critical thinking skills is achieved through the process of forcing them to determine the keywords which most clearly represent the topic they want to research. This is the key to all bibliographic research, not only in the CIF, but also in every library catalog, every commercial index and most other resources in print. The success of this process was measured by a survey comparing the number of keywords to the number of subject headings in the same records.

Measurement of the amount of staff time involved was readily achieved. Librarian time required to

assign subject headings using the hand production system was already known. Clerical time needed to produce and file cards was measured. What remained was to measure the time required to produce greatly expanded records electronically. With staff time at a premium, it was essential that inputting complete tables of content along with the bibliographic record must not take significantly longer than the hand production of cards containing only bibliographic records with broad subject headings. Any increase in the time required would result in delays, meaning that students would not have access to the latest information on their topics. As it turned out, the time factor was very close.

The author took full responsibility for establishing the conversion procedure. Responsibilities included, but were not limited to, experimentation with various formats, creation of electronic templates, development of input procedures and creation of a procedure manual, training of librarians in search techniques, training and supervision of support staff who would do the actual inputting, and proofreading.

CHAPTER II

Research and Solution Strategy

Literature Survey

The author took on the task of gathering information on computer programs for indexing specially produced library files. Library Literature was searched, and a few pertinent articles were found. An ERIC search yielded many citations, only one of which was useful for the intended purpose. By far the most usable information came from PACS-L, the electronic library computer conference.

This age of electronic information has provided library users with more than they can possibly use. Online catalogs, CD-ROMs and other retrieval systems make it possible for a student to enter a seemingly simple search and end up with hundreds of citations, few of which will be of any use for the immediate assignment. Keyword access makes it possible for the beginning researcher to suffer information overload on a grand scale (Jaffe, Sanders, 1991). Far from being elated by this wealth of information, the hapless and inexperienced freshman or sophomore is frustrated by the vast number of citations which require examination, many of which will have no bearing on the assignment.

The problem is, online searches of library catalogs are restricted to bibliographic citations consisting of author, title, place of publication, publisher, date of publication and any assigned Library of Congress (LC) subject headings. Keyword searching will pick up any bibliographic entry containing the specific words entered, but what is needed is access to the CONTENTS of books and other printed materials.

Currently, LC subject headings only indicate the overall subject content of a document (DeHart, 1989). The purpose of LC subject headings is to provide a summary of contents, rather than specifics. These broad headings offer virtually no bibliographic access to the specific topics contained in a given work. While a chapter or section of a book might suffice, there is virtually no way of determining, from online catalog information, the real contents of a book. Subject access provided is to the whole, rather than to its parts (Frantz, 1990). Numerous useful items of information are never found, because they are cataloged under such a broad context (Pinn, 1988). Subject cataloging under the present system is inadequate. Keyword searching utilizing the same system produces far too many irrelevant citations (Lipow, 1983).

Access to the tables of contents would greatly enhance catalog records, improve search effectiveness and help the student decide whether the book addresses the needs of a particular assignment. Words from the contents page could be indexed as keywords (DeHart, 1989). While LC headings are too broad, and full-text searching produces far too many irrelevant citations, a limited-text bibliographic database could be created from tables of contents, which usually tend to be sufficiently descriptive (Tompkins, 1990). Such a database would be the electronic equivalent of systematic browsing, that time-honored practice of going to a section of the library then methodically taking each book in hand to examine its blurb, table of contents and index (Mann, 1987).

While the use of keywords in tables of contents creates the hazards of uncontrolled vocabulary, it would also improve user access and retrieval effectiveness in ways not previously available (Bates, 1989). While the typical keyword index is created for very large collections, using mainframes, some libraries are now experimenting with the same concept, using personal computers, to index smaller but intensively used collections (Jaeger, 1988).

At this point, programming is still experimental and tentative. Whatever a library has by way of personal computers and programs can be explored. Some measure of success has been reported using Macintosh Plus with Microsoft Works 1.0 (Foster, 1990), an IBM PC with Datafile (Jaeger, 1988), HyperCard (Bantz, Kambitsch, Cavrak, 1991), dBASE II (Harrington, 1991), and Compuserve (LaBorie, 1991), to name just a few. A major goal of keyword indexing is to provide the capability of natural language searching (Kambitsch, Kreitz, 1991). Natural language, as opposed to formal, broad subject headings, approaches the topic using commonly known terminology and syntax.

Solution Strategy

What seemed clear from available literature and from responses to the author's inquiry on PACS-L, was a trend toward libraries using available technology to create small, very user-friendly indexes to the contents of specified types of materials. The reason for this was the same as the reason that libraries, long before the era of electronic information access, started indexes and formalized subject headings in the first place - to help readers locate specific materials at the time of need. From ancient times, the holdings

of libraries have been organized into broad subject areas. But, as material in broad subject areas became more comprehensive, classification became more difficult. Gaining access to specific bits of information became as time-consuming as it was before the advent of subject cataloging. Readers were faced with the problem of too much material, too little specific retrieval.

Small computers with powerful programs offered the potential for providing faster access to specific pieces of information contained within broad subject publications. Librarians began experimenting with a variety of programs, and achieving some measure of success. No one program has provided the ultimate key to full realization of the goal of full natural language information retrieval, so the field was very much open to experimentation.

When this project began, there were no fully developed commercially available keyword indexing programs capable of handling academic library inhouse files with retrieval strategies suitable for beginning researchers. The college library owned HyperCard, which was the program most often mentioned in the literature as giving good results for keyword retrieval. The task was to develop it for community

college beginning level research. To accomplish this, it was necessary to completely recreate the bibliographic records of the the existing CIF on HyperCard.

"Recreate" did not mean duplicate. Library of Congress (LC) subject headings had to give way to natural language terms. The original idea was to physically examine each publication listed in the CIF, to determine and assign the most appropriate keywords in context (KWIC). It was thought that KWIC would supplement LC subject headings, to the extent that they approximated natural language. A core sample was taken from the CIF, including entries from each of the five resources. The items were retrieved, and every possible keyword in the title and table of contents was considered. Great care was taken to make sure that words chosen conformed to standard indexing, by comparing the words with entries in the three-volume Library of Congress Subject Headings.

The exercise proved to be extremely time-consuming, a factor which would affect how quickly material would become available to students. In addition, there were frustrations due to LC's penchant for obfuscation and formality. The decision was made to simply "go with

the publishers" entering bibliographic information and tables of contents without judgement.

It was a good move in two major ways. First, librarian involvement was reduced to proofreading of data entered. Proofreading took much less time than selection of formal subject headings or keywords, and allowed the librarian to proofread for only the natural language which occurred in the publications. Second, the time saved contributed toward more quickly getting the information out to the public area.

One of the hazards of extensive subject card files was the problem of multiple cards necessitated by multiple subject headings. Unlike commercially produced catalog cards, most of the entries in the inhouse card file lacked tracings; there was no way to know for sure how many cards were on file for any given item. Considering keywords as possible subject headings, it quickly became evident that searching the card file for all potential cards relating to a given item would not be practical. Instead, it was decided to "start from scratch" with original materials. For instance, the CIF listed holdings for Congressional Digest from 1979 to the present. Clerical staff had only to begin there, pulling the 1979 bound volume of

Congressional Digest, photocopying the contents pages of each issue, then inputting data from the photocopies. The original articles remained available to users, and the photocopy was later used to assist in proofreading input data.

Macintosh computers were already used for specified library assignments. Program floppy disks could be checked out upon presentation of a current student ID card. There was some concern early on that the file would grow beyond a single disk, which might have a negative effect on keyword searching. It became evident that would not be a problem. Experience indicated that the disk space required to add full tables of content was easily offset by the fact that only one bibliographic entry per item was needed, unlike the card file where a separate bibliographic entry was needed for each subject heading. The library's practice of keeping files current assurred that older materials would be removed when disk space became a problem.

CHAPTER V

Method

The idea for the project was conceived the summer of 1990, when the college librarians commented that there should be some way of converting the Controversial Issues File (CIF) into some kind of electronic format, just as the main card catalog had been converted from card file to electronic file. The author was working toward an advanced degree at the time, and was considering various options for a project which would both benefit the library and satisfy degree requirements. Later in the year, it was agreed that the author should be the one to pursue the conversion.

During fall semester 1990, the author, an IBM user, took a class to learn Macintosh. During the last two class periods in December, the class was introduced to HyperCard. The librarians were already aware of HyperCard, and believed it would be useful for the conversion project, but none of the librarians were HyperCard users. At the author's request, the instructor put together brief directions for creating a stack, and led the author through the procedure to devise a card. It seemed easy enough at the time, with help close to hand.

During December and January, parameters were discussed with other librarians, an extensive but disappointing literature search was conducted, and the proposal was written. All agreed that a simple conversion without change would be a waste of time. The conversion had to be a definite improvement, in the form of keyword searching capabilities. Actually, what was really desired was full Boolean capabilities. Unfortunately, both available literature and advice from experienced HyperCard users indicated that Boolean was beyond HyperCard, but keyword was within its range. Since keyword searching would be a great improvement over the existing card file containing only formal LC subject headings, it was agreed that the project would still be worth doing.

To begin, a core sample of 25 bibliographic records was pulled from the CIF. The sample included entries from all five resources: Congressional Digest, Editorial Research Reports, Opposing Viewpoints, Reference Shelf, and Taking Sides. Care was taken that each of the entries was unrelated to any of the others, in terms of subject content, and that there were no duplicate records. The CIF is strictly a subject catalog, and the subject entries chosen spanned the

alphabet. The cards were photocopied, then returned to the file.

Using the photocopies, an alphabetical list of subject headings was compiled. Some of the entries had tracings, indicating additional subject headings under which the bibliographic records were filed. The sample of 25 item records yielded 63 subject headings. That is, in most cases, a student could locate information by using any one of two or three different subject approaches.

Next, each of the books or periodicals in the sample was physically pulled from the shelves. The title page and table of contents for each item were photocopied. Significant words were marked, and listed on a form created for determining keywords (Appendix A:44). To ensure that keywords represented as accurately as possible standard terminology for the contents of the item, cross-reference checks were conducted, using The Readers' Guide to Periodical Literature, Library of Congress Subject Headings, the Magazine Index, and the National Newspaper Index, all of which combine formal subject headings with natural language.

This procedure served a number of purposes. First, it established the beginning of an authority file,

which researchers could later use as a printed index to the electronic CIF. Second, by timing how long it took to establish keywords for each item, a comparison could be made with the already-known time required to determine the more restrictive formal subject headings. An unexpected benefit was the discovery that every keyword chosen was in the Library of Congress Subject Headings, and nearly every word was also included in the other three authorities. This meant that the validity of significant words chosen as keywords was virtually assurred. Precious time could be saved by checking only the LC Subject Headings.

But what if a researcher searched for a topic using a keyword not recognized by the author? For instance, a survey of students using the CIF (Appendix B:46) indicated that they tended to start with either very broad subject headings, such as ABORTION, or very specific cases, such as Webster v. the State of Missouri. Without the benefit of the Survey, the author would have listed Webster v. the State of Missouri under the letter W, but not also under M for Missouri. The student who was searching for that information did not remember Webster, but did recall that the case originated in Missouri, so looked for information under Missouri as a subject heading in the

card file; it was not listed. One item under the subject ABORTION in the CIF does list Webster v. the State of Missouri in its table of contents.

The only way to ensure discovery of any word chosen by a user, if contained in the record, would be to enter the full table of contents. With electronic keyword searching and full table of contents, the student could have entered Missouri, and that item record would have been retrieved. Table of contents as part of the bibliographic record has been 'The Impossible Dream' of librarians for a long time. Contents take considerable time to enter, and would consume exhorbitant amounts of drawer space in a card catalog. The author surmised that contents would also take up a lot of disk space in an electronic catalog.

Computer experts on campus were consulted. It was determined that, considering the present size of the file - 1500+ records - and its rate of growth of five per month, plus new series editions at intervals of several years, the amount of disk space used on the 40 megabite hard drive would be negligible. The decision was made to enter full table of contents with each bibliographic record.

Having made that decision, the next logical step became obvious: drop the idea of a keyword index. The

purpose of an index is to determine whether there are any references to a specific word. Indexes are not generally browsed. Rather, the user thinks of a word and looks for that word in the index. With HyperCard, the same word can be entered under the Find command, and the program takes the user through the file, showing every occurrence of that word, which is the same as keyword searching in the online public access catalog (OPAC). The difference between the OPAC catalog and the CIF is that the OPAC does not have tables of content, but the CIF soon will. By adding tables of content to the CIF, the user sees the chosen word in exactly the context intended by the publisher.

A new background card was created. The first background card, made in class before the project began, proved inadequate in practice. It contained a Subject field, which was eliminated. While subject headings utilize well-known words, they do not always use those words in the way that ordinary people use them; the words are familiar but the arrangement is not. The core sample showed that, with only one exception, the keyword in each subject heading also appeared somewhere in the title and/or the table of contents. A user searching for the keyword would find the record in any case. The one exception was a

subject heading, relating to education, which used bueraucratic language unknown to this educator, while the table of contents of that record used the well-known keyword. Also, tables of content are often quite extensive, extending well beyond one screen. In addition to dropping the Subject field, Scrolling was added to the Contents field. Scrolling is a HyperCard feature which extends one field indefinitely without losing sight of the other fields. New buttons were added, and all fields were restructured to visually enhance the overall appearance of the card.

It should be noted here that changing the background card was quite an experience for the author. It was stated earlier that HyperCard was introduced in a class during the final two class periods of a Macintosh course, and that a background was created in class with the instructor, in effect, leading the author by the hand through the process. At that time, brief written instructions were also handed out. As it turned out, two class periods and brief instructions were not quite enough to make this individual an expert in creating a HyperCard stack.

Considerable literature about HyperCard was available, including material which came with the program. Most of it assumed the reader already knew

more than the author knew at the time. None of it contained the details which could have helped to create a stack and understand the steps in a way which could be applied to future new stacks of this nature. The author had to start from scratch and learn HyperCard. This meant temporarily suspending work on the project. Based on this experience, the author wrote a procedure manual for creating a HyperCard stack. This manual was not a textbook, nor did it contain comprehensive coverage of all that HyperCard had to offer. Rather, it fully documented the steps required to create one stack: the Controversial Issues File. The procedures as written were full of the sort of details which the author longed for but did not find while searching available literature. After that, work on the project was resumed.

From the beginning, the purpose of the conversion project was and remains to make the content of materials more accessible to students. To this end, it was hoped that the time required for converting each record would not greatly exceed the time presently required to produce and file cards. It was surmised that adding tables of content would take longer than the present system of assigning only a few very broad subject headings.

It was already estimated that assigning subject headings, creating and filing cards took an average of eighteen to twenty minutes per item. Practice with the core sample proved the estimate was very close. Note here that Reference Shelf is exceptionally in-depth, and Taking Sides covers a large number of broadly related social issues per volume. In terms of the number of individual cards produced for the card file, real time was adjusted to average per card time.

Time Required to Assign Subject Headings, Produce and File Cards (Appendix C:47).

Congressional Digest	21 mins
Editorial Research Reports	19 mins
Opposing Viewpoints	20 mins
Reference Shelf	36 mins (18 mins)
Taking Sides	228 mins (12 mins)

Cards also had to be covered and filed. These procedures took an average of 40 seconds per card.

The time required to determine subject headings, hand-produce and file cards took an average of nineteen minutes. The time required to determine subject

headings, produce and file a set of cards using Quick Card took an average of twenty-one minutes. With both procedures, what took the most time was determining the appropriate LC subject headings. By changing to a practice of simply entering bibliographic and contents information, it was no longer necessary to determine subject headings. But, how much more time would it take to enter tables of content? For comparison, it was necessary to determine how much time was needed to produce electronic entries. Records were kept, detailing the amount of time required to input data.

Time Required to Enter Data, Including Full Table of Contents (Appendix D:48).

Congressional Digest	18 mins
Editorial Research Reports	12 mins
Opposing Viewpoints	11 mins
Reference Shelf Series	37 mins
Taking Sides Series	120 mins

While it appeared obvious that electronic data entry without subject headings but with tables of content took less staff time, proofreading had to be factored in. With a card file, the information would

be found even if a card were slightly misfiled or a word slightly misspelled, but a misspelled word on an electronic database was lost forever to keyword searching.

Time Required to Underline Keywords and Proofread Electronic Records (Appendix E:49).

Congressional Digest	3 mins
Editorial Research Reports	4 mins
Opposing Viewpoints Series	3 mins
Reference Shelf Series	8 mins
Taking Sides Series	33 mins

Therefore, more realistic figures for the time involved in electronic production, in real time, were:

Congressional Digest	21 mins
Editorial Research Reports	16 mins
Opposing Viewpoints Series	14 mins
Reference Shelf Series	45 mins
Taking Sides Series	153 mins

While it was originally thought that providing additional data would require much more time, that has not been the case. Although the times required for electronic production compare very favorably with card file production, keyword capabilities are a plus factor. Searching by keyword requires no knowledge of formal subject headings, and students are already accustomed to keyword searches on a number of other library databases. Since titles and tables of content both tend to be descriptive, the number of possible keywords is much greater than the number of possible subject headings. The sample of 25 item records yielded 63 subject headings. In the same sample, 219 keywords were identified. Put another way, library users will have greatly improved odds of locating material on their topics, no matter what word approach is taken. The increased benefit to students is well worth the time involved in conversion.

Once the stack was established, data entry became a very simple procedure. It was thought that both librarians and clerical staff would require extensive training for the conversion. As it turned out, librarians needed only minimal training for purposes of proofreading. A prerequisite for clerical staff was knowledge of word processing on a Macintosh, and the

introductory lessons which came with the HyperCard program. As a number of clerks were regular Macintosh users, the first prerequisite was already met. It was not necessary to be a HyperCard expert to input data once the stack had been established, so the training required was minimal.

Using the library's existing HyperCard program, the process has begun to convert the Controversial Issues File (CIF) from cards to electronic format utilizing keyword retrieval capabilities. The author has personally performed all functions necessary to establish procedures for completion of the conversion project. To this end, that included all professional and clerical tasks needed to establish complete procedures. A procedure manual has been written, in accordance with library policies.

The card file remains intact throughout the coversion, which is in process, so students have the usual access to it. In addition to writing procedures for the clerk to input data, the author has also written a search procedure for information retrieval on the electronic CIF. To this end, the procedures for inputting included experimenting with operators which would make simple search procedures possible. There are no uniform library search procedures for

electronic information retrieval. The library presently has seven electronic information retrieval databases, and every one requires different search strategies, so students quickly learn to look for printed directions. While many students are underprepared for college level research, they are, by and large, familiar with computers, and seem to have little trouble adapting from one system to another.

Timetable of Activities During Implementation.

Week 1: Worked on assigning keywords as subject headings.

Week 2: Experimented with new backgrounds. Computer people loaded HyperCard 2.0. Had been working with 1.2.5.

Week 3: Learned that 2.0 did not say much to 1.2.5. Time out to learn more about using 2.0.

Week 4: Spent the week just beginning to really get acquainted with HyperCard.

Week 5: Continued learning HyperCard.

Week 6: Took some keywords from annotations of chapters, to supplement some chapter headings. Worked some on format of bibligraphic information. Started writing instructions for beginning data entry clerk. In so doing, helped to refine technique. Input entire contents of one Taking Sides. Looked into the possibility of getting an optical character recognition (OCR) scanner.

Week 7: Established the amount of time required to input records from each of the five resources. Demonstration and discussion with Mentor. Pondered whether it was necessary to create a printed keyword list. Worked on more detailed instructions.

Week 8: Created a totally new stack design, based on discoveries over the past few weeks. Bad news: software for the OCR scanner has been lost. Good news: Computer Lab thinks there may be a way to upgrade the stack for Boolean searching. Discovered some duplication of entries, caused by the lack of tracings in the card file.

Week 9: Computer Lab upgraded the 1.2.5 stack to 2.0, providing some needed enhancements. Learned how to script arrow buttons, thus eliminating many steps involved in linking cards. Still in the process of learning HyperCard features.

Week 10: Computer Lab confirmed what literature indicated: OCR scanning does not eliminate the need for proofreading. It is still worth trying, and software has been ordered. Manual input continued. Finally learned the difference between scipting background cards and scripting non-background cards.

Week 11: After much experimentation and learning, finally started writing detailed procedures for creating this stack. Other libraries will need to know.

Week 12: Continued working on detailed procedures.

Week 13: Completed procedure manual for creating the stack. Worked on instructions for staff who will be entering data.

Week 14: Decided order of input should be reverse chronological. Eliminated working from the card file in favor of working off photocopies of title and content pages. After writing directions for students to use the stack, had a student try it. Modified directions.

Week 15: Completed input procedures for clerical staff. Training and supervision commenced. Now that procedures have been establisheded, completion time for full conversion is estimated to be approximately nine months.

CHAPTER IV

Results

This practicum had two prime objectives. First, devise a way to improve the Controversial Issues File (CIF) by converting it from a subject-based card file to a keyword-based electronic file, which would greatly increase students' chances of locating specific information. Second, on a per item basis, because of limitations on staff time, the conversion and subsequent upkeep must not greatly exceed the amount of time required to produce the original card file.

To compare the number of subject headings per item in the card file with the number of keywords which could be accessed electronically, a core sample of 25 cards was examined. The sample included records from all five sources, and spanned the alphabet. Care was taken to avoid any duplication of records or subject headings. The 25 records had a total of 63 subject headings. This meant that a student searching for information on one of these topics could locate the desired record if the two or three exact subject headings were known.

Next, title and content pages for each of the sample records were photocopied. Keywords were marked and counted. The 25 records had a total of 219

keywords. A student looking for information on one of these topics could, on average, locate the record using any one of nine words. No knowledge of formal subject headings is necessary to search for keywords.

While the core sample clearly demonstrated that electronic keyword searching greatly increased access to information, it remained to be seen whether the time required to produce electronic records would exceed the College Library's available staff time. Time comparisons for the steps involved with card production and electronic production were made, using four methods of measurement. The first measurement, the Time Required to Establish Keywords (Appendix A:44) was originally created as Time Required to Establish Subject Headings. Based on the assumption that keywords, like formal subject headings, should conform to standardized indexing, it seemed reasonable that keyword assignments should follow the same procedure. Upon learning that all but one of the keywords was contained in a legitimate formal subject heading, it was decided that checking every keyword against the Library of Congress Subject Headings would not be necessary. Figures for establishing both subject headings and keywords were useful in making total time comparisons.

Before proceeding with other production measurements, one class of 31 students, assigned to use the CIF for research papers, served as a survey sample. The Controversial Issues File Survey (Appendix B:46) was used to determine the effectiveness of subject headings versus keywords. It became evident that, if subject headings were simple, natural words, such as ABORTION, then the records were as easily located using either subject headings or keywords. But, if the formal subject heading was not commonly known, such as MILITARY SERVICE, COMPULSORY for the better known keyword DRAFT, then keyword searching really proved its value.

Getting back to the mechanics of inputting data versus subject card production, the Time Required to Assign Subject Headings, Produce and File Cards (Appendix C:47) was measured. Card production was done in two different ways. For the weekly publication, Editorial Research Reports, cards were written by hand; all others were input on computer using the Quick Card program. Hand production and filing took an average of nineteen minutes. Quick Card production and filing took an average of twenty-one minutes.

The next measurement was the Time Required to Enter Data, Including Full Table of Contents (Appendix D:48).

With an electronic file, no cards were produced, so no filing was needed. Therefore, real time rather than per card time had to be compared. Also, with electronic files, spelling is of the essence. Therefore, it was necessary to learn the Time Required to Underline Keywords and Proofread Electronic Files (Appendix E:49). When all factors were considered, the comparison between time required to produce card files and time required to produce electronic files was very close.

	Number of Minutes	
	CARDS	ELECTRONIC
Congressional Digest	21	21
Editorial Research Reports	19	16
Opposing Viewpoints	20	14
Reference Shelf	36	45
Taking Sides	228	153

When this project was first conceived, it was hoped that conversion could be accomplished in not much more time than it took to create and build the original card file. Extra measurements were taken for Reference Shelf, but it persisted in being the only one to require more time for electronic production than hand

production. The extra time required to input Reference Shelf data is more than made up for by the reduced time required to input all other records. It was originally estimated that complete conversion of the CIF would take two academic years, or four semesters. Based on these figures, cautious hope has risen to genuine optimism that conversion will be accomplished in one academic year.

CHAPTER V

Recommendations

Once conversion has been completed, copies on floppy diskettes will be made available to other libraries in the community college district. Since all academic libraries create inhouse files for ongoing assignments, and all District libraries have HyperCard and Macintosh computers, printed procedure manuals will be duplicated and distributed. Feedback will be encouraged, and any suggested improvements will be considered.

Instructions for students to use the file are very simple and are on the file. In addition, printed instructions will be available. Students here are accustomed to using a variety of electronic files, and seldom have any difficulty. Instructions are basically the same as for keyword searching on the OPAC.

Input via an optical character recognition (OCR) scanner is still considered an option. Product evaluations indicate that some characters can be misread by scanners, and therefore incorrectly reproduced (Diehl, Eglowstein, 1991), making careful proofreading a necessity. Another sticking point is the variety of font and size on title and content pages, which could confuse the scanner. Most scanner

programs are capable of converting a number of fonts to one uniform font, but take extra time to accomplish the frequent changes (Meng, 1991). Even so, the Computer Lab has ordered a new OCR scanner and program, and experimentation will take place while conversion is in progress.

The one major disappointment with the conversion project is HyperCard's lack of Boolean search capabilities. It can search keywords much like the OPAC, but Boolean searching would be even more beneficial to students. Some of the campus computer experts think it can be programmed for Boolean searching, and an especially capable Computer Lab student employee is presently working on it.

Upon completion of the conversion, the author plans to publish an article in a library or education periodical. Also, a paper will be submitted for presentation at the annual State Library Association conference. Hopefully, both an article and a conference paper will be accepted within a year of conversion.

References Cited

NOTE: A number of references are identified as "Response to PACS-L inquiry." The Public Access Computer Systems Forum (PACS-L) is an international computer conference that deals with all computer systems which libraries make available to their patrons. PACS-L is accessed through the College's electronic mail system via Bitnet. The aurhor first signed on to the forum with an inquiry on Jan. 6, 1991.

Annual Report of Phoenix College Library, 1989/90.

Bantz, David A. Response to PACS-L inquiry, Jan. 8, 1991.

Bates, Marcia J. "Rethinking subject cataloging in the online environment," Library Resources and Technical Services, 33 (4), Oct., 1989.

Cavrak, Steve. Response to PACS-L inquiry, Jan. 7, 1991.

Day, Mary. "MCCCD 45th day certified enrollment," District Office, Maricopa County Community College District, Phoenix, Arizona, Fall, 1990.

DeHart, Florence E. and Richard Reitsma. "Subject searching and tables of contents in single-work titles," Technical Services Quarterly, Vol. 7(1), 1989.

Diehl, Stanford and Howard Eglowstein. "Tame the paper tiger," Byte, April, 1991.

Foster, Constance L. "Subject access to periodicals: a main-Mac combo," College and Research Libraries News, February, 1990.

Frantz, Paul. "A gaping black hole in the bibliographic universe," American Libraries, July/August, 1990.

Harrington, Jim. Response to PACS-L inquiry, Jan. 16, 1991.

Jaeger, Sheila E. "Microcomputer keyword indexing in government documents," Library Software Review v.7, Jan/Feb., 1988

Jaffe, Lee. Response to PACS-L inquiry, Jan. 4, 1991.

Kambitsch, Tim. Response to PACS-L inquiry, Jan. 17, 1991.

Kreitz, Pat. Response to PACS-L inquiry, Jan. 8, 1991.

LaBorie, Tim. Response to PACS-L inquiry, Jan. 13, 1991.

Lipow, Anne Grodzins. "Practical considerations of the current capabilities of subject access in online public catalogs," Library Resources & Technical Services, January/March, 1983.

Lynch, Anne. "Project Jefferson and the development of research skills," Reference Services Review, Fall, 1989.

Mann, Thomas. A guide to library research methods. New York: Oxford University Press, 1987.

Meng, Brita. "Text without typing," Macworld, October, 1990.

Pinn, Ken. "The library in the computer," School Librarian, November, 1988

Sanders, Thomas. Response to PACS-L inquiry, Jan. 8, 1991.

Teitelbaum, Sheldon. "USC's new teaching library," USC Trojan Family, September, 1990.

Tompkins, Philip. "Community college library and information technology: Planning and our transitional environment." Handout provided by Tompkins at his seminar Nov. 6, 1990.

U.S. Census Bureau, Phoenix District Office, Phoenix, Arizona. Information obtained by telephone July, 1990.

van der Laan, Hans. Independent inquiry on PACS-L Jan. 4, 1991.

Appendix A

Time Required to Establish Keywords.

1. Clock time started this item:

2. Title of periodical or series.

3. Title of article or section.

4. LC subject headings already assigned.

5. Keywords taken from item. (Mark T if in title.)

6. Readers' Guide to Periodical Literature

   List keywords from #5 not appearing in R.G. (Do not list here if keywords are listed with "see" references.)

   List "see" references, if any, for keywords from #5 listed in R.G.

7. Magazine Index

   List keywords from #5 not appearing in M.I. (Do not list here if keywords are listed with "see" references.)

   List "see" references, if any, for keywords from #5 listed in M.I.

8. National Newspaper Index

List keywords from #5 not appearing in N.N.I. (Do not list here if keywords are listed with "see" references.)

List "see" references, if any, for keywords from #5 listed in N.N.I.

9. Library of Congress Subject Headings

List keywords from #5 not appearing in L.C.S.H.. (Do not list here if keywords are listed with "see" references.)

List "see" references, if any, for keywords from #5 listed in L.C.S.H.

10. Final list of keywords for this item.

11. Time finished:

Appendix B

Controversial Issues File Survey

The Library is trying to improve the Controversial Issues File. Please help us to help you by filling out this form.

1. Please state briefly what you were looking for.

2. Did you look in the OPAC (the online catalog) before looking in the Controversial Issues File?

Yes_____ No_____

3. If yes, please list subjects or keywords tried.

4. Please list subjects or keywords you looked for in the Controversial Issues File.

THANK YOU FOR COMPLETING THIS SURVEY. THE INFORMATION WILL BE USED TO IMPROVE THE CONTROVERSIAL ISSUES FILE.

PLEASE LEAVE THIS COMPLETED FORM AT THE REFERENCE DESK

Appendix C

Time Required to Assign Subject Headings, Produce & File Cards.

1. Card production start time:

2. Card production finish time:

3. Filing start time:

4. Filing finish time:

APPENDIX D

Time Required to Enter Data, Including Full Table of Contents.

Title of periodical or series

Data entry start time

Data entry finish time

Total elapsed minutes

APPENDIX E

Time Required to Underline Keywords and Proofread Electronic Records.

Title of periodical or series

Keyword start time

Keyword finish time

Proofreading start time

Proofreading finish time

Total elapsed minutes